Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 33, Num. 1 (2026) 69-86

RESEARCH ARTICLE

Machine Learning and Explainable Artificial Intelligence in Carbon Nanotubes Predictions

Aprendizado de Máquina e Inteligência Artificial Explicável em Previsões de Nanotubos de Carbono

Moacir Guedes Oliveira[1], Carolina Paula Almeida[2], Sandra Mara Guse Scós Venske[1,2*]

Abstract: Nanotechnology is an area that can be applied in all fields of science and nanomaterials, such as Carbon Nanotubes (CNT), have been the subject of studies in order to reduce costs, time for experiments and development. CNTs have diverse applications because they are highly rigid, have high strength, and have low density. Because of this, the evaluation of the properties of carbon nanotubes and identification of their coordinates are important fields of research for science. Many real problems predictions, including those on carbon nanotubes, are Multi-Target Regression (MTR) tasks, where multiple continuous target variables are predicted. Machine learning techniques are well suited to address these types of problem. However, the literature reports approaches composed of complex models, which are manually configured and do not present any level of explainability. The goal of this work is to use an Automated Machine Learning (AutoML) technique, for which 10 classic machine learning algorithms are available. The proposal, named aASL-MTR, uses ensemble learning and AutoML to choose an ensemble of algorithms and their hyperparameters to handle MTR problems. Three CNT data sets were used in the experiments, with 3 and 4 targets, to predict the coordinates and properties of the carbon nanotubes. We also tested our approach on a set of 18 real-world data sets, with different numbers of targets, to evaluate its performance. In addition, an explainable artificial intelligence method was applied to CNT problems. Statistical tests were used and the aASL-MTR algorithm was compared favorably with classical algorithms and also with the literature.
Keywords: Carbon Nanotubes — Multi-target Regression — Machine Learning — Automated Machine Learning — Explainable Artificial Intelligence

Resumo: A nanotecnologia é uma área que pode ser aplicada em todos os campos da ciência, e os nanomateriais, como os Nanotubos de Carbono (NTC), têm sido objeto de estudos com o objetivo de reduzir custos, tempo de experimentos e desenvolvimento. Os NTCs possuem diversas aplicações por serem altamente rígidos, apresentarem alta resistência e baixa densidade. Por isso, a avaliação das propriedades dos nanotubos de carbono e a identificação de suas coordenadas são campos importantes de pesquisa para a ciência. Muitas previsões de problemas reais, incluindo aquelas relacionadas aos nanotubos de carbono, são tarefas de regressão multi-alvo, nas quais múltiplas variáveis-alvo contínuas são previstas. Técnicas de aprendizado de máquina são bem adequadas para lidar com esse tipo de problema. No entanto, a literatura apresenta abordagens compostas por modelos complexos, que são configurados manualmente e não apresentam nenhum nível de explicabilidade. O objetivo deste trabalho é utilizar uma técnica de Aprendizado de Máquina Automatizado (AutoML), para a qual 10 algoritmos clássicos de aprendizado de máquina estão disponíveis. A proposta, denominada aASL-MTR, utiliza aprendizado por *ensemble* e AutoML para escolher um conjunto de algoritmos e seus hiperparâmetros para lidar com problemas de regressão multi-alvo. Três conjuntos de dados de NTCs foram utilizados nos experimentos, com 3 e 4 variáveis-alvo, para prever as coordenadas e propriedades dos nanotubos de carbono. A abordagem também foi testada em um conjunto de 18 bases de dados do mundo real, com diferentes números de variáveis-alvo, para avaliar seu desempenho. Além disso, um método de inteligência artificial explicável foi aplicado aos problemas de NTC. Testes estatísticos foram utilizados e o algoritmo aASL-MTR foi comparado favoravelmente com algoritmos clássicos e também com propostas da literatura.
Palavras-Chave: Nanotubos de carbono — Regressão multi-alvo — Aprendizado de Máquina — Aprendizado de Máquina Automatizado — Inteligência Artificial Explicável

[1] *Postgraduate Program in Nanosciences and Biosciences*, Universidade Estadual do Centro-Oeste (UNICENTRO), Brasil
[2] *Department of Computer Science*, Universidade Estadual do Centro-Oeste (UNICENTRO), Brasil
*__Corresponding author__: ssvenske@unicentro.br
DOI: http://dx.doi.org/10.22456/2175-2745.146905 • **Received:** 14/04/2025 • **Accepted:** 26/12/2025

1. Introduction

Carbon nanotubes (CNT) are cylindrical graphene structures that are one of the research targets in the nanosciences area [1]. Nanoscience is basically a combination of three areas: science, engineering, and technology, to study matter and processes at the nanoscale. The nanoscale is based on the nanometer, a unit of length equivalent to one billionth (10^{-9}) of a meter. Nanotechnology can be applied in all fields of science and nanomaterials have been studied to reduce costs, time for experiments, and development [2]. CNTs have practical application in several areas due to their high rigidity, high strength, and low density. Therefore, evaluating its properties and identifying its coordinates are important research fields for science [1].

Recently, Machine Learning (ML) techniques have been of considerable interest in exploring predictions involving carbon nanotubes [1]. ML is a field of Artificial Intelligence (AI) that aims to replicate the fundamental aspects of human learning, such as observation, repetition, and the analysis of multiple examples to identify underlying patterns. In ML, the pattern recognition process is based on probabilistic estimates. The precision with which ML models can represent various systems has attracted the attention of scientists [3].

Many predictions related to real problems, including those related to carbon nanotubes, are Multi-Target Regression (MTR) tasks. Multi-target regression involves predicting multiple continuous target variables simultaneously using a common set of input variables. Despite receiving comparatively less attention from the machine learning community, multi-target regression plays a crucial role in numerous real-world applications [4]. Thus, in addition to carbon nanotube data sets, the proposed method is tested on 18 other multi-target data sets benchmarks, making predictions in several application areas, such as water quality, air ticket value, and river network flow.

Today, many companies of different sizes have made use of Automated Machine Learning (AutoML) to obtain high-quality machine learning models in an easier way. AutoML encompasses techniques that automate the selection, training, and tuning of machine learning models. Sometimes, ensemble learning, which involves training multiple learners simultaneously to solve a problem, can be combined with AutoML.

In this work, we use an AutoML method, Auto Sklearn (ASL), which makes use of ensemble learning and performs algorithm selection and hyperparameter tuning. ASL has been adapted in order to work with a large set of ML algorithms, adapted for the multi-target regression task. This proposal is named aASL-MTR (adapted Auto Sklearn for Multi-Target Regression). The classic algorithms available in the ensemble to be used for a multi-target regression task were: Automatic Relevance Determination (ARD) [5], Stochastic Gradient Descent (SGD) [6], Decision Tree (DT) [7], K-Nearest Neighbors (KNN) [8], Support Vector Machines (SVM) [9], Gaussian Process (GP) [10], Random Forest (RF) [11], Extra Trees (ET) [12], Ada Boost (AB) [13] and Gradient Boosting (GB) [14]. This pool of algorithms was selected in order to combine a wide range of regressors, as some of the algorithms are ensemble learning-based and non-ensemble-based.

In order to better understand the results and output of the machine learning algorithm, an eXplainable Artificial Intelligence (XAI) method was applied to problems related to CNT. XAI is used to describe an AI model, its expected impact, and possible biases.

The main contributions of this work are: i) improvement in the prediction of coordinates and mechanical properties of carbon nanotubes, since there are few works applying machine learning techniques for these considered data sets; ii) how approaches using ensembles of classical algorithms can help in dealing with prediction of coordinates and mechanical properties of carbon nanotubes, since the literature uses algorithms of artificial neural networks and deep learning in the approach; iii) study the prediction of mechanical properties of carbon nanotubes using a combination of attributes not considered in the literature; iv) use of an explainable artificial intelligence approach for problems related to carbon nanotubes; v) adapting algorithms to work with multi-target regression in an AutoML framework, specifically Auto sklearn, since ASL was underexplored for the MTR task; vi) performance analysis of the proposed method against the literature for different data sets for multi-target regression, covering different amounts of samples, attributes, and targets.

This paper is organized as follows. Section 2 presents the background for the proposed work, addressing concepts of machine learning, carbon nanotubes, and related work. In Section 3 the proposed approach is described. The results are presented and analyzed in Section 4. Finally, Section 5 presents conclusions and future directions.

2. Background

This section initially presents machine learning concepts (Subsection 2.1), information on the importance of studying carbon nanotubes (Subsection 2.2), and some related works (Subsection 2.3).

2.1 Machine Learning

Machine learning (ML) is a branch of Artificial Intelligence (AI) that focuses on developing algorithms and models capable of automatically learning and making predictions or decisions from data, without being explicitly programmed [3]. Some ML concepts related to this work will be briefly discussed below.

2.1.1 Multi-Target Regression

The regression task in machine learning involves predicting a continuous or numerical value based on input data. We seek to find a functional relationship between the input variables and the continuous output variable. MTR is a task in which the machine learning algorithm predicts multiple real-valued variables simultaneously. Formally, it can be defined as finding a function $f : X \rightarrow Y$ such that f maximizes c, where [15]:

- X is an input space, with tuples of dimension d, containing values of primitive data types, i.e., $\forall x_i \in X$, where $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})$;
- Y is an output (target) space, with tuples of dimension t, containing real values, i.e., $\forall y_i \in Y$, $y_i = (y_{i1}, y_{i2}, \ldots, y_{it})$, where $y_{ik} \in \mathbb{R}$ and $1 \leq k \leq t$;
- S is a set of examples where each example is a pair of tuples from the input and the output space, i.e., $S = \{(x_i, y_i), |, x_i \in X, y_i \in Y, 1 \leq i \leq N\}$ and N is the number of examples in S, $(N = |S|)$;
- c is a quality criterion which rewards models with high predictive accuracy and low complexity.

In this work, the function f is represented by an ensemble of classic algorithms, managed by an AutoML algorithm. The function f was also considered to test the classic algorithms individually.

2.1.2 Ensemble Learning

The ensemble methods [16] involve training multiple learners simultaneously to address a common problem. Unlike traditional learning approaches that aim to construct a single learner based on training data, ensemble methods focus on creating a collection of learners and merging their outputs [3, 17]. An ensemble built from different machine learning techniques can be classified as heterogeneous, while with the use of a single technique, the ensemble is classified as homogeneous. The construction of an ensemble takes place through the use of some Ensemble Method (EM) [18, 3]. Ensemble learning methods can be broadly classified into three main categories [16, 3, 17]: bagging, stacking, and boosting.

Bagging (bootstrap aggregation) involves parallel training of ML techniques on different subsets of the training data through bootstrapping and combining their predictions through averaging or voting. Stacking involves training multiple ML techniques and then training a meta-model to combine their predictions. Boosting focuses on sequentially training ML techniques that correct the mistakes made by previous models, resulting in a strong ensemble.

Ensemble learning has been proven to be effective in various domains, including classification, regression, and anomaly detection. It has been successfully applied in real-world scenarios to improve prediction accuracy, generalization, and robustness of machine learning models [3].

2.1.3 Automated Machine Learning

AutoML is a research field that aims to automatically select, compose, and parameterize machine learning models to optimize the performance of a given task that involves a data set [19]. The autoML pipeline has some typical components. The first component consists of a data preparation step that involves loading and cleaning data for use in the system, as well as transformations, normalization, or encoding. The next component involves the step of selecting the features that will be used in creating the model. The final components involve the iterative steps of building, training, optimizing, validating, and selecting a machine learning algorithm to use on a given problem [20].

Auto Sklearn (ASL) is a tool that uses auto machine learning concepts and aims to find the best options, considering a set of machine learning algorithms. For its operation, ASL uses the Bayesian optimization method [21]. ASL can be configured to be homogeneous or heterogeneous. Figure 1 summarizes the ASL pipeline.

For optimization using the Bayesian method, ASL included two steps in the AutoML pipeline. The first step is meta-learning (to initialize the Bayesian optimization procedure), which results in improved system efficiency. The second step included is an automated ensemble construction system, which allows using all the regressors (or classifiers) found by Bayesian optimization.

The Bayesian optimizer pursues the best models combined with their best hyperparameters, evaluating the performance of each model, and selecting the best hyperparameters for the next iteration. The process is repeated iteratively until a stopping criterion is reached. ASL performs the creation of an ensemble from the models considered the most promising during Bayesian optimization. For this, the ensemble selection technique from Caruana *et al* [22] was used as EM, which in short starts with an empty ensemble and iteratively adds the best performing model until reaching a limit size.

2.1.4 Explainable Artificial Intelligence

eXplainable Artificial Intelligence (XAI) provides methods that aim to produce machine learning models that present information to balance issues involving interpretability versus accuracy. XAI aims to build white/gray box ML models that are interpretable by design while achieving high accuracy or black box models that exhibit a minimal level of interpretability when white/gray box models cannot achieve an admissible level of accuracy [23].

SHAP[1] (SHapley Additive exPlanations) is a tool that uses the principles of XAI. Its primary purpose is to assess the significance of each input attribute and quantify its impact on the prediction result [24]. Shapley values, derived from game theory, encompass the degrees of significance of the attributes assigned [25]. Essentially, the Shapley value is the expected average marginal contribution of an attribute after all possible combinations have been considered.

2.2 Carbon Nanotubes

Carbon nanotubes have emerged with the aim of solving problems caused by the miniaturization of materials. They are cylindrical graphene structures [26, 27]. The two main structural parameters of a carbon nanotube are the diameter and chiral angle, which define its structure (also called chirality).

There are two classes of carbon nanotubes: single-walled carbon nanotubes (SWCNTs) and multi-walled carbon nanotubes (MWCNTs). SWCNTs and MWCNTs have different

[1] https://pypi.org/project/shap/

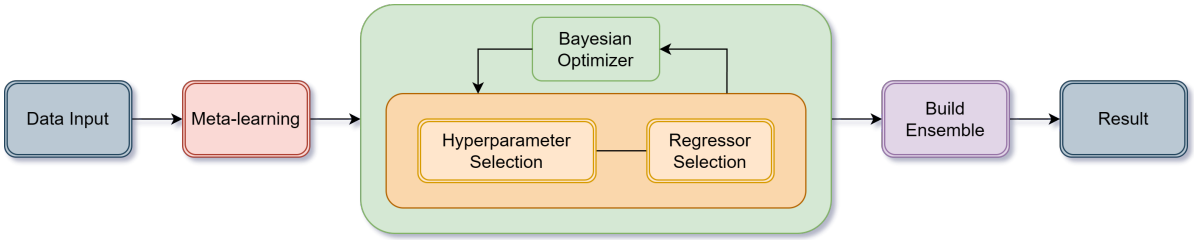

Figure 1. Schematic overview of Auto-sklearn. Adapted from [21].

physical properties because of their structural differences. In SWCNTs, the nanotube wall is formed by only one layer of graphene. They are graphene sheets rolled up to form cylinders. MWCNTs consist of a concentric arrangement of the SWCNTs, forming several layers of graphene rolled into cylinders [28].

SWCNTs are used in industrial paints, car tires, structural materials, adhesives, lubricants, transparent conductive films, cables, among others [28]. Potential applications of MWCNTs are additives in polymers, catalysts, electron field emitters for cathode ray lighting elements, flat panel display, lithium-battery anodes, drug delivery, among others [28].

In this work, two SWCNT data sets were considered, both of which are related to multi-target regression task. CASTEP-CNC (CASTEP software using Carbon Nanotube Coordinates)[1] is a data set that contains information that allows predicting the atomic coordinates of carbon nanotubes. The SWCNT-Canadija[2] data set contains basic data on the configurations and corresponding mechanical properties of single-walled carbon nanotubes up to 4 nm in diameter. Information about attributes and targets of these data sets is presented in Section 4.1.

2.3 Related Works

Our work is related to three main areas: the problem of predictions related to carbon nanotubes, general multi-target regression task, and ensemble learning. Initially, in this section, the works that approach the two multi-target regression problems related to CNT are presented. Then, different works related to other multi-target regression tasks and ensemble learning are also listed.

Carbon nanotubes have applications in various fields. A review of the use of machine learning with nanotechnology and nanomaterials was conducted, focusing on carbon nanotubes [1]. Many of the works address the use of machine learning for the classification/prediction of different thermal, mechanical, and electrical properties of carbon nanotubes.

Only two works [26, 29] were found in the literature that deal with the problem of predicting the atomic coordinates of carbon nanotubes, both using the CASTEP-CNC data set, consisting of 5 attributes and 3 targets. Four neural network models are used to perform coordinate prediction: Feed-forward neural networks (FFNNs), Function Fitting Neural Network (FITNET), Cascade-Forward Neural Network (CFNN), and Generalized Regression Neural Network (GRNN) [26]. The objective was to reduce the time needed to obtain the final coordinates, as the usual mathematical methods can take days to calculate. The FFNN and FITNET models obtained better results. The metrics used to measure performance were R, MSE (Mean Squared Error), MAE (Mean Absolute Error), and SSE (Error Sum of Squares). Artificial neural networks are also applied to predict the coordinates of carbon nanotubes [27]. In this work, the FFNN model was also used along with two other models: Residual Neural Network (ResNet) and Convolutional Neural Network (CNN). The results for the CASTEP-CNC data set were specified in terms of precision, considering a relationship between the actual results and the predictions of the networks.

Regarding the prediction of carbon nanotube properties, an artificial neural network was proposed in conjunction with deep learning to predict four different mechanical properties of carbon nanotubes [29]. The SWCNT-Canadija data set was used to predict 4 targets considering the R^2 metric.

There are different studies related to multi-target regression and ensemble learning tasks [30, 31, 32, 33, 34, 35, 36, 37, 38, 4, 39]

In another work a rule learning method is proposed for the multi-target regression task, called FIRE-ROS (Fitted Rule Ensembles with Random Output Selections) [30]. The method extends the rule learning method FIRE and learns an ensemble of Predictive Clustering Trees (PCTs) with random output selections (ROS). For the experiments, 16 benchmark data sets containing between 6 and 35 targets with different number of instances per data set were used. Model performance is measured using the aRRMSE (Average Relative Root Mean Squared Error) metric.

Ten machine learning algorithms were applied to a data set containing herbaceous biomass samples to quantitatively determine biomass feedstock compositions (one target) [34]. The algorithms were as follows: SVRLin (Support Vector Regression with Linear kernel), SVRRad (Support Vector Regression with Radial kernel), LASSO (Least Absolute Shrinkage and Selection operator), RIDGE (RIDge REgression), ENET (Elastic NET regression), PLS (Partial Least Squares regression), RPART (Recursive Partitioning and Regression Trees), RF (Random Forest), GBM (Gradient Boosting Machine) and GP (Gaussian Process). The algorithms were tested individually and later combined to form a stacked ensemble.

[1] Available at https://archive.ics.uci.edu/ml/data sets/Carbon+Nanotubes
[2] Available at https://data.mendeley.com/data sets/vwd34rvphw

The results in terms of RMSEP (Root Mean Square Error of Prediction) and R^2 metrics were independently verified between the algorithms and then compared with the ensemble result. The stacked ensemble obtained the best results compared to the other machine learning techniques used.

Four artificial neural networks are used to optimize the result of solar and wind forecasts from data obtained in Spain and Brazil (forecast time series) [35]. The neural network models used were Backpropagation Multilayer Perceptron, RBF (Radial Bases Function), CFBP (Cascade Forward Back Propagation), and SOM (Self-Organizing Map). After individual applications of neural networks, with the results obtained in terms of RMSE, R, MAE and MAPE (Mean Absolute Percentage Error), an ensemble is performed using Ridge Regression in order to obtain a performance increase. Using the ensemble improved performance compared to using the algorithms in isolation.

A method called MTR-TSF that deals with multi-target regression tasks by learning Target Specific Features (TSF) is proposed [36]. In MTR-TSF, the targets are linked to the specific attributes of each data set. The proposed method uses the regression tree boosting method (CART-boosting). Eighteen data sets with the number of targets varying between 2 and 16 are used in the experiments, and the results are evaluated with the aRRMSE metric.

In another work the dependencies that targets in multi-target problems have on each other are studied [37]. The authors proposed two ensemble methods called SST (Stacked Single-Target) and ERC (Ensemble of Regressor Chains) and tested them on 18 multi-target regression data sets, with the number of targets varying between 2 and 16. The RRMSE (Relative Root Mean Squared Error) metric is used.

Three SVM-based models were proposed to deal with multitarget regression problems [38]. The first builds single-target regression models for each target, called the support vector regression (SVR). The second builds an ensemble using the previously generated models, a model called SVRRC, a multi-target ensemble of randomly chained SVR models. The third calculates the relationships between the targets and assembles a model called SVRCC (SVR Correlation Chains). The experiments are carried out in 24 data sets (with the number of targets varying between 2 and 16) and showed that the SVRCC model obtains promising results in relation to the other compared models, considering the aRRMSE, MSE, aCC (average Correlation Coefficient) and aRMSE (average Root Mean Squared Error) metrics.

Structural modeling is proposed where the relationships of the targets are modeled in a nonlinear way [4]. The approach, denoted as NOSL (Non-linear Output Structure Learning), was tested on 6 data sets (with data sets having 2 to 16 targets) in terms of aRRMSE metric.

A multi-target regression algorithm called Multilayer Multi-target Regression (MMR), built from the Elastic Net algorithm, is presented, it also uses the concept of multilayers in a similar way to neural networks [39]. Experiments to evaluate the performance of the MMR were performed on 18 data sets, which have from 2 to 16 targets. The aRRMSE metric is used.

Our proposal aASL-MTR connects the approach of using ensemble learning and the automated machine learning method for the treatment of multi-target regression problems. We used 21 real-world data sets that contain targets ranging from 2 to 16 and that encompass a wide variety of MTR tasks. Also, we present the results in terms of 5 metrics: MSE, SSE, MAE, R and aRRMSE. Approaches for predicting the coordinates and mechanical properties of carbon nanotubes are still rare, and these problems are little explored computationally. Finally, the use of an explainable artificial intelligence technique contributes to understanding in the construction of machine learning models.

3. Proposed approach

The general scheme of our proposal is presented in Figure 2.

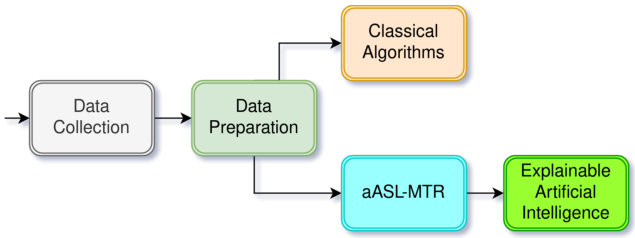

Figure 2. Scheme of the proposed approach. The proposed algorithm - aASL-MTR is represented in the Auto Machine Learning step.

3.1 Data Preparation

In the first stage, the data is collected (reading the data sets), followed by preprocessing. In the preprocessing of all data sets, blank numerical values were replaced by the median.

Transformation of symbolic data to numeric was done using the *OrdinalEncoder* encoder[3].

The normalization technique by patterning (defined in Equation 1) was used. All values to be normalized ($V_{Current}$) are subtracted by the average value of their attributes (μ) and the resulting value is divided by the standard deviation of the attributes (σ). With this, a new data set is generated, with values of mean 0 and variance 1 [40].

$$V_{New} = \frac{V_{Current} - \mu}{\sigma} \tag{1}$$

3.2 Models Generation

In the next step of the scheme, the algorithms that will generate the models are used: considering classic algorithms or considering the automatic machine learning algorithm.

A set of 10 classical algorithms was considered with their hyperparameters adjusted by a technique that chooses the

[3] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OrdinalEncoder.html

best parameters from a given set of parameters in a grid[4]. The algorithms used were as follows: ARD, SGD, DT, KNN, SVM, GP, RF, ET, AB, and GB. This pool of algorithms allows the automated machine learning algorithm to explore a broad set of solutions, including ensemble learning-based and non-ensemble-based regressors.

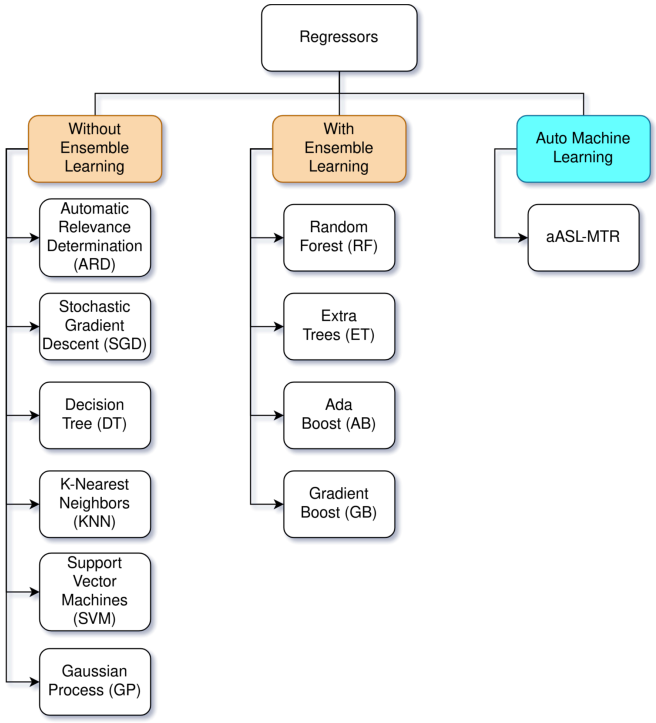

Figure 3. Algorithms used in the proposed scheme. The proposed algorithm - aASL-MTR is represented in the Auto Machine Learning category.

The automated machine learning step in Figure 2 represents the application of the Auto-Sklearn (ASL) toolkit that frees the user from algorithm selection and hyperparameter tuning. It takes advantage of recent advantages in Bayesian optimization, meta-learning, and ensemble construction [21]. Our proposal aASL-MTR adapts the ASL to use the 10 classic mentioned algorithms, adjusted and modified to work as multi-target regression. No feature selection method was applied. Figure 3 shows all algorithms used, classified as machine learning algorithms with and without ensemble learning. It is important to note that the pool of algorithms available for aASL-MTR is made up of all the classical algorithms considered.

3.3 Generated Models Evaluation

The evaluation stage consists of verifying the performance of the previously created models. There are some metrics options in the literature for this purpose. For this work, the following metrics were used: Mean Squared Error (MSE), Sum of Squared Error (SSE), Mean Absolute Error (MAE), Coefficient of Multiple Correlation (R), and Average Relative Root Mean Squared Error (aRRMSE).

Cross-validation was used to assess performance through different splits performed in the training set, where each split configuration is used to train an ML algorithm [41, 40].

An explainable artificial intelligence technique (SHAP) is applied to the models generated by aASL-MTR. The development was carried out through implementation in Python[5] programming language, along with the open source tools Scikit-learn[6], Auto Sklearn[7] and SHAP.

4. Results

In this section, the results of the experiments performed are presented and discussed. The section is divided into four subsections: in subsection 4.1 the data sets used in the experiments are presented; in subsection 4.2 the test methodology is explained, including the metrics and statistical tests used during the analysis; finally, in subsections 4.3 and 4.4 the results for carbon nanotubes and general MTR are presented.

4.1 Considered data sets

In this work, 21 data sets were used, detailed in Table 1. In the first stage of the results, 3 multi-target data sets related to carbon nanotubes were considered, with the number of targets varying between 3 and 4. In the second stage of the results, 18 multi-target data sets commonly used in the literature were considered[8], with a number of targets ranging from 2 to 16.

The CASTEP-CNC data set is for predicting the coordinates of carbon nanotubes; it contains 3 targets that refer to the atomic coordinate numbers of the nanotubes themselves. The data set has five attributes, which are the chiral indices n and m; initial atomic coordinates u, v and w (CASTEP randomly generated u, v and w parameters of the initial atomic coordinates of all carbon atoms, respectively). The three targets are: calculated atomic coordinates u', v' and w' (CASTEP calculated u', v' and w' parameters of the atomic coordinates of all carbon atoms, respectively). These parameters represent the fractional atomic coordinates of all carbon atoms within the crystallographic unit cell of the carbon nanotube structure, as calculated by CASTEP. In this context, u', v', and w' specify the position of each atom as a fraction of the unit cell's dimensions along its respective crystallographic axes.

The SWCNT-Canadija data set was originally used in the work of Canadija [29]. The author used two chiral indices (n and m) from the data set to predict four different mechanical properties of carbon nanotubes (Young's modulus initial, initial Poisson's ratio, fracture strain, and ultimate tensile strength (UTS). Young's modulus is a mechanical property that measures the tensile or compressive stiffness of the material when the force is applied in the longitudinal direction [42].

[4]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[5]https://www.python.org/

[6]https://scikit-learn.org/

[7]https://www.AutoML.org/AutoML-for-x/tabular-data/au to-sklearn/

[8]Available at https://www.openml.org/

Table 1. Description of the 21 data sets used in experiments.

data set (*S*)	Size (*N*)	Attrib. N. (*X*)	Targ. N. (*Y*)
	1^o Stage		
CASTEP-CNC	10721	5	3
SWCNT2-Canadija	818	2	4
SWCNT4-Canadija	818	4	4
	2^o Stage		
ANDRO	49	30	6
ATP1D	337	411	6
ATP7D	296	411	6
EDM	154	16	2
ENB	768	8	2
JURA	359	15	3
OES10	403	298	16
OES97	334	263	16
OSALES	639	413	12
RF1	9125	64	8
RF2	9125	576	8
SCM1D	9803	280	16
SCM20D	8966	61	16
SCPF	1137	23	3
SF1	323	10	3
SF2	1066	10	3
SLUMP	103	7	3
WQ	1060	16	14

The initial Poisson ratio is a measure of the Poisson effect, which includes the deformation of a material in perpendicular directions [43]. Fracture strain is the elongation at break (ductility measurement) of a carbon nanotube [44]. UTS is the maximum resistance of the carbon nanotube to fracture.

The SWCNT-Canadian data set also contains two additional attributes (initial diameter and length of carbon nanotubes) that were not used in the original work [29], because the author concluded that the chirality of carbon nanotubes has a large impact on the mechanical properties to be predicted. However, in this work, we decide to consider all the attributes in order to explore the data set as much as possible. In the experiments, we consider two different versions of the data set: i) SWCNT2-Canadija that contains only the two chirality indices as attributes (to make possible the comparison with the original work [29]), and ii) SWCNT4-Canadija that contains four attributes (two chiral indices, initial diameter and initial length).

The other 18 data sets were used in the experiments to test aASL-MTR on a classic set of benchmarks for MTR data sets. The ANDRO (Andromeda) data set has six targets corresponding to the prediction of the water quality of a gulf in Greece. The available samples were collected using underwater sensors [45].

The data sets ATP1D and ATP7D (Air Ticket Price) correspond to the price of airline tickets. Each sample contains 6 targets and information from different flight departures that are used to predict the value of future air tickets on the next day (ATP1D) or the minimum possible value in the next 7 days (ATP7D) [46].

The EDM (Electrical Discharge Machining) data set addresses a 2-target regression problem that aims to make a machine control the value of two variables, as a human would control [47].

The ENB (Energy Building) data set also works with two targets, which have information related to heating and cooling loads for building construction [48].

The JURA data set has 3 targets consisting of measurements of concentrations of seven different heavy metals that have been collected in the soil of the Jura Mountains [49].

The OES10 and OES97 (Occupational Employment Survey) data sets contain information from occupational employment surveys from the United States Bureau of Labor Statistics. The data were obtained in 1997 (OES97) and 2010 (OES10), both have 16 targets [37].

The OSALES data set is a preprocessed version of the data set used in Kaggle's "Online Product Sales" competition. Each sample contains 12 targets that correspond to the number of products sold online every month in a year [50]. In another competition called Kaggle's "See Click Predict Fix" the SCPF data set was used, where each sample contains 3 targets that are related to the number of views, clicks, and comments on a complaint mechanism in the United States [51]. Both the OSALES and SCPF data sets had blank cells. To avoid this, the median value of each column was used to fill these fields.

Data sets RF1 and RF2 (River flow) contain information to predict the flow of the river network for the next 48 hours. RF1 contains information from 8 different locations on the Mississippi River. RF2 adds an extension by including information related to the precipitation forecast of the 8 locations [37]. Both have 8 targets and had to go through an additional treatment to fill in the blank cells using the median of corresponding column.

The SCM1D and SCM20D (Supply Chain Management) data sets have 16 targets and information collected from the "Trading Agent Competition" tournament. Each sample corresponds to one day of the tournament, the data sets aim to predict the average ticket price for the next day (SCM1D) or the next 20 days (SCM20D) of the tournament [37].

The SF1 and SF2 (Solar Flare) data sets correspond to the number of solar flares observed in a 24-hour period. SF1 contains data from 1969 while SF2 from 1978 [52] both have 3 targets. In these data sets, the symbolic fields were transformed into numerical ones.

The SLUMP data set (Concrete Slump) has 3 concrete properties as target [53].

Finally, the WQ (Water Quality) data set presents targets related to 14 parameters of the physicochemical quality of river water in Slovenia [54].

4.2 Experiments Methodology

In this work, the following metrics were used during the analysis of the results: MSE, SSE, MAE, R and aRRMSE. These metrics are defined by Equations 2 to 6, where n is the number of instances, Y_i corresponds to the values of the targets of the instances, $\hat{Y}$ are the results of the predictions and $\bar{Y}$ represents the average values of the targets of the instances. All metrics measure between 0 and 1, and with the exception of the R metric, smaller values are preferred.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2}$$

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{4}$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{5}$$

$$aRRMSE = \frac{1}{n}\sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{6}$$

In all experiments, the K-fold cross-validation method is used with k value as 10. All classic algorithms were parameterized using a grid search algorithm [19]. The different value ranges used for the grid search to choose the most appropriate value are shown in Table 2 (the ranges were chosen as suggested by Shahhosseini *et al.* [55]). aASL-MTR eliminates the need for manual configuration by autonomously performing hyperparameter optimization to construct an ensemble. Nevertheless, it allows the specification of the algorithm's time limit and the metric employed for its internal optimization. Across all runs, a fixed time limit of 60 minutes was enforced, and the RMSE was utilized for optimization purposes.

The experiments are divided into two stages. The first stage analyzes the results related to carbon nanotubes, and the second stage treats the results involving the other multi-target regression data sets. Statistical tests were performed at both stages of the experiments. In the first stage, the Friedman statistical test was used, while the ANOVA statistical test was used in the second stage. The ANOVA and Friedman statistical tests were used, according to the recommendations of Demšar [56]. All tests were performed with a significance level of 0.05 and the MSE metric was considered.

The tests perform the validation of the null hypothesis H_0 that all methods are equivalent, if it is rejected, the Bonferroni-Dunn post-hoc test [57] was used. Calculate the Critical Distance (CD) between the methods, which is the minimum distance that two methods need to have in terms of ranking to be considered significantly different. The CD calculation is given by Equation 7 [56].

Table 2. Classical algorithms and their hyperparameter settings [55].

Algorithm	Hyperparameters	Values
ARD	alpha1 alpha2	$10^{range(-5,1)}$ [1] $10^{range(-5,1)}$
SGD	alpha l1_ratio	$10^{range(-5,1)}$ $10^{range(-5,1)}$
DT	max_depth	range(3, 100)[2]
KNN	n_neighbor	linspace(2, 20, 10)[3]
SVM	C kernel	linspace(0.01, 5, 20) {linear, poly, rbf}
GP	alpha	$10^{range(-10,-4)}$
RF	n_estimators max_depth	{100, 200, 500} range(3, 10)
ET	n_estimators max_depth	{100, 200, 500} range(3, 10)
AB	n_estimators learning_rate	{100, 200, 500} linspace(0.5, 2, 20)
GB	n_estimators learning_rate	{100, 200, 500} linspace(0.5, 2, 20)

[1] Values between 10^{-5} and 10^{1}, incremented by 1.
[2] Values between 3 and 100, incremented by 1.
[3] 10 values linearly spaced between 2 and 20.

$$CD = q\alpha\sqrt{\frac{M(M+1)}{6N}} \tag{7}$$

such that M is the number of compared methods, N is the number of data sets, and $q\alpha$ is the critical value which has fixed values predefined by Demšar [56]. Regarding how the comparison is performed, the Bonferroni-Dunn test designates one of the methods as the control method and then compares its performance against the others (one against all). When the diagram is constructed from the Bonferroni-Dunn test, the range of critical differences is marked under the main line, and all the methods within the marking are not considered to be significantly different. In addition, the control method is highlighted [56].

The principles of explainable artificial intelligence were put into practice using the SHAP tool. aASL-MTR leverages diverse ensembles, resulting in more intricate models compared to conventional algorithms. For this reason, the SHAP analysis was used only in the models generated from aASL-MTR in the data sets from the first stage of the tests.

4.3 First Stage Results - Carbon Nanotubes

In this stage, the results of the experiments carried out with the data sets of carbon nanotubes named CASTEP-CNC, SWCNT2-Canadija and SWCNT4-Canadija are presented.

Table 3 presents the results for the CASTEP-CNC data set, the data set that predicts the atomic coordinates of the carbon nanotubes. MSE, SSE, MAE and R metrics are used to evaluate the results. The literature for the CASTEP-CNC data set does not use the aRRMSE metric, so it was not considered.

Analyzing Table 3, it is possible to observe that aASL-MTR obtained a superior performance compared to the classic algorithms in all metrics considered. The ARD and SGD algorithms performed well, while the SVM algorithm achieved the worst results. With these tests, it is concluded that aASL-MTR is the most suitable algorithm for predicting the atomic coordinates of carbon nanotubes. Statistical tests using the ANOVA test were applied to the 11 algorithms compared considering the MSE metric. The input for the tests were the values obtained in each of the 10 folds from the cross-validation. At this stage, the null hypothesis was not rejected, which implies that all compared methods are statistically equivalent.

To our knowledge, the work of Acı and Avcı [26] is the only in the literature that uses the CASTEP-CNC data set and uses four different neural networks: Feedforward Neural Network (FFNN), Function Fitting Neural Network (FITNET), Cascade-Forward Neural Network (CFNN), and Generalized Regression Neural Network (GRNN). The comparison of the aASL-MTR results is shown in Table 4.

aASL-MTR was better evaluated considering the SSE metric, and the results of the literature [26] were superior in the other metrics. It should be noted that the work in the literature [26] considers neural networks specifically designed to solve the problem of predicting carbon nanotubes, which differs from aASL-MTR, which, due to its adaptive nature, can be applied to other problems. In addition, the proposed approach releases the user from the full configuration of the ensemble, allowing non-expert users in machine learning to obtain high-quality models.

Table 5 presents the results for the SWCNT2-Canadija data set, which predicts the mechanical properties of carbon nanotubes. In this case, aASL-MTR was far superior to the other algorithms considering all metrics. The RF algorithm was the second best result, and the SVM algorithm was the worst algorithm to predict the mechanical properties of CNTs. Statistical tests using the ANOVA test were applied to the SWCNT2-Canadija data set considering the 11 methods for the MSE metric. The input for the tests were the values obtained in each of the 10 folds from the cross-validation. The null hypothesis was rejected, pointing to the existence of a statistical difference between the compared algorithms. The Bonferroni-Dunn post-hoc test was applied and the critical difference diagrams are shown in Figure 4, considering aASL-MTR as the baseline algorithm.

The Bonferroni-Dunn post-hoc test points that aASL-MTR is rank 1 and is statistically equivalent to the RF, GB and DT algorithms. This group of three algorithms is statistically superior to the eight others compared according to the statistical test. It is worth noting that of the 3 algorithms statistically equivalent to aASL-MTR, only one is not based on ensemble (DT).

Table 6 shows the results for the SWCNT4-Canadija data set. Again, aASL-MTR outperforms the classical algorithms. The GP algorithm was the second best result, and again the SVM was the worst algorithm. Statistical tests using the ANOVA test were also applied to the SWCNT4-Canadija data set, also considering the 11 methods for the MSE metric. The input for the tests were also the values obtained in each of the 10 folds from the cross-validation. The null hypothesis was rejected, pointing to the existence of a statistical difference between the compared algorithms. The Bonferroni-Dunn post-hoc test was applied and the critical difference diagrams are shown in Figure 5, considering aASL-MTR as the baseline algorithm.

The Bonferroni-Dunn post-hoc test points that aASL-MTR is rank 1 and is statistically equivalent to the GP, ET, RF and GB algorithms. These algorithms are statistically superior to the others according to the test. Among the 4 algorithms statistically equivalent to aASL-MTR, only GP is not based on the ensemble.

For comparison with the literature for the SWCNT-Canadija data sets, Canadija's work was used [29], since it is the only work that uses this data set, in which a Deep Neural Network (DNN) is applied. The comparison of the aASL-MTR results is shown in Table 7.

It can be seen that the results for the SWCNT4-Canadija data set were superior to those obtained by SWCNT2-Canadija in all metrics considered. Therefore, it can be concluded that using the 4 attributes of the data set (n, m, initial diameter, and initial length) is better than using only 2 attributes (n and m). Compared to Canadija' work [29], which uses only 2 attributes, the R metric points out that aASL-MTR outperformed the approach proposed by Canadija [29]. It should be noted that deep learning techniques were used in Canadija's work [29], which are more robust and computationally expensive compared to those applied in this work.

For the three data sets involving the prediction of coordinates and properties of carbon nanotubes, aASL-MTR showed good results.

4.3.1 Analysis of the Generated Ensembles in Carbon Nanotubes data sets

This section explores the composition of the ensembles generated by the proposed approach for carbon nanotube data sets. Figures 6 and 7 present, for each of the CNT data sets, the relationship between the size of the set generated by aASL-MTR and the diversity of ML techniques that make up it. For the three data sets, Figure 6 shows that the techniques are repeated in the composition of the ensemble, with different configurations of their hyper parameters. The ensemble generated for SWCNT2-Canadija is the most populous and the most heterogeneous, followed by CASTEP-CNC and SWCNT4-

Table 3. Results of first stage experiments on the CASTEP-CNC data set with different metrics. The best values are highlighted in dark gray and the second best are in light gray.

Algorithm	MSE	SSE	MAE	R
ARD	0.000086542409	0.01852814	0.001996821	0.9994894
SGD	0.000086641779	0.01856142	0.002048057	0.9994888
DT	0.001141752359	0.23339465	0.023101653	0.9932436
KNN	0.001232433167	0.26434840	0.025443566	0.9926687
SVM	0.003234388693	0.69377119	0.048215015	0.9806258
GP	0.000774550651	0.16619233	0.009406665	0.9803550
RF	0.000337507639	0.07205917	0.011817558	0.9980051
ET	0.000257479497	0.05463849	0.009448841	0.9984776
AB	0.000922206783	0.18873919	0.019520367	0.9944896
GB	0.000166608512	0.03562638	0.003396928	0.9990262
aASL-MTR	0.000074060236	0.01591950	0.001793987	0.9995647

Table 4. Comparison of the aASL-MTR results with the literature for the CASTEP-CNC data set. The best aRRMSE values are highlighted in dark gray.

Algorithm	MSE	SSE	MAE	R
FFNN [26]	0.000006679809	0.02003943	0.001536561	0.9999597
FITNET [26]	0.000006625736	0.01987721	0.001478840	0.9999602
CFNN [26]	0.000010766950	0.03230084	0.002130755	0.9999354
GRNN [26]	0.078235160000	77.7360000	0.224040300	0.9412479
aASL-MTR	0.000074060236	0.01591950	0.001793987	0.9995647

Canadija. Figure 7 shows the total number of appearances of each of the ML techniques in the ensembles of the three CNT data sets and also in how many ensembles (data sets) each technique has. The AB and SVM techniques are the ones that appear in greater quantity in the composition of two ensembles. This fact demonstrates the ability of the aASL-MTR in tuning hyperparameters, since when the tuning was done by grid search, the SVM presented the worst performance. Among the ML techniques available to aASL-MTR for the creation of the ensemble some do not appear in the ensemble of any of the three data sets, they are: SGD, GB, DT, KNN, and GP. The following algorithms with their parameters adjusted by grid search are statistically equivalent to aASL-MTR for CNT data sets: GP, ET, RF, GB (SWCNT2-Canadija); RF, GB, and DT (SWCNT4-Canadija) and for the CASTEP-CNC data set, all algorithms are equivalent to each other. Among these equivalents, only GB, DT and GP do not appear in any CNT ensemble.

4.3.2 Explainability in Carbon Nanotube data sets

The explainability of the models generated by aASL-MTR was explored for the CASTEP-CNC and SWCNT4-Canadija data sets, using the SHAP tool. The results are presented graphically and demonstrate how much each attribute cooperates for the values of the coordinate predictions or mechanical properties of the carbon nanotubes. Figures 8 and 9 show the results for the CASTEP-CNC and SWCNT4-Canadija data sets, respectively.

Figure 8 shows the attributes that contributed the most to the aASL-MTR model considering the CASTEP-CNC data set. It also provides information about the specific impact of the specific attribute on the prediction of each target. The three atomic coordinates are the features that most influence the prediction. Thus, the initial atomic coordinate attribute u contributed the most to the prediction of the calculated atomic coordinate u' and contributed slightly to the prediction of the calculated atomic coordinate v'. The initial atomic coordinate of the attribute w contributed the most to the prediction of the calculated atomic coordinate w' and contributed slightly to the prediction of the calculated atomic coordinate v'. The prediction of atomic coordinates v', u', and w' hardly uses chiral indices of the attributes.

Figure 9 shows the attributes that contributed the most to the aASL-MTR model considering the SWCNT-Canadija data set. The initial diameter is recognized by SHAP as the most important attribute. This attribute is the one that most influences the prediction of the Young's modulus initial. This target even uses all the attributes to be predicted. The prediction of the UTS target is also influenced by the four attributes. The Initial Poisson's ratio and Fracture Strain targets have a low influence on the attributes (which was omitted in the illustration due to the scale of the graph).

In a detailed study [29], was identified that the chirality of nanotubes (attributes m and n) was identified to have a great impact on the prediction of the mechanical properties of nanotubes. This was confirmed in this analysis of explainability in the construction of the machine learning model for the SWCNT4-Canadija data set. However, it can be seen that the initial diameter attribute also had a great influence on the model.

Table 5. Results of first stage experiments on the SWCNT2-Canadija data set with different metrics. The best aRRMSE values are highlighted in dark gray and the second best are in light gray.

Algorithm	MSE	SSE	MAE	R
ARD	48.251043669483	7913.17116179	2.69473794	0.7950159
SGD	48.251612124478	7913.26438841	2.69675760	0.7950380
DT	33.089499448626	5426.67790957	2.51179709	0.8063651
KNN	33.891359494590	5558.18295711	2.549898616	0.7986300
SVM	49.272449189248	8080.68166703	2.65906292	0.4458818
GP	34.907710514891	5724.86452444	2.57804288	0.8154876
RF	32.216345155167	5283.48060544	2.46117942	0.8125158
ET	33.804785301100	5543.98478938	2.50555503	0.8121810
AB	37.141678528492	6091.23527867	2.69288742	0.7999103
GB	32.628265139412	5351.03548286	2.48109443	0.8074713
aASL-MTR	7.297442022820	1196.780491742	1.06527849	0.9793250

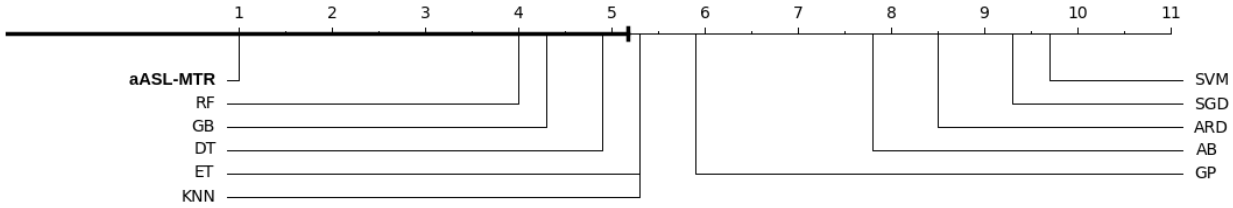

Figure 4. Graphical representation of the Bonferroni-Dunn post-hoc test displaying the average rank of the MSE for the 11 algorithms considered for the SWCNT2-Canadija data set. The critical difference was equal to 2.813 and aASL-MTR was the baseline algorithm.

4.4 Second Stage Results - Multi-target data sets

In this step, the results of the experiments carried out with the 18 multi-target data sets are presented (Table 1). The adopted metric was the aRRMSE, which is used in the literature to evaluate these data sets.

Table 8 displays the results of the experiments for the 18 data sets considering the 10 classical algorithms and aASL-MTR. In terms of absolute results, aASL-MTR was the best algorithm for 10 data sets, taking the second-best result for 4 additional data sets. ET was the second overall best algorithm, with the best results for 4 data sets. Friedman's statistical test was applied and indicated that there is a statistical difference between the compared algorithms. The Bonferroni-Dunn post-hoc test was applied and the critical difference diagrams are shown in Figure 10, considering aASL-MTR as the baseline algorithm.

The Bonferroni-Dunn post hoc test points that aASL-MTR is rank 1 and is statistically equivalent to the ET and RF algorithms, both based on ensembles. This group of algorithms is statistically superior to the other 8 compared according to the considered statistical test. For comparison of the literature, the following works were used (presented in Section 2.3): SST [37], ERC [37], MMR [39], SVRCC [38], MTR-TSF [36], NOSL [4], and FIRE-ROS [30].

Table 9 displays the results of the experiments for the 18 data sets considering the 10 algorithms from the literature, and aASL-MTR. In terms of absolute results, aASL-MTR was the best algorithm for 6 data sets. The MTR-TSF algorithm outperformed all algorithms in 9 data sets.

Friedman's statistical test showed a significant difference between the algorithms analyzed. For the test, six algorithms (SST, ERC, MMR, SVRCC, MTR-TSF, and aASL-MTR) were considered, in which the original papers presented complete results for all 18 data sets. The Bonferroni-Dunn post-hoc test was applied and the critical difference diagrams are shown in Figure 11, considering aASL-MTR as the baseline algorithm.

It can be seen in the graph of Figure 11 that the algorithms aASL-MTR, MTR-TSF and MMR are statistically equivalent and superior to the other 3 algorithms compared.

Some of the data sets used in this step have a large amount of attributes, making the use of the XAI concepts a very complex task. For this reason, SHAP was not used in this step and this is a planned future work.

4.4.1 Analysis of the Generated Ensembles in general MTR data sets

This section explores the composition of the ensembles generated by aASL-MTR for the multi-target regression data sets. Figures 12 and 13 present, for each of the MTR data sets, the relationship between the size of the set (the total number of ML techniques) and the diversity of techniques that make up it. For all data sets, Figure 12 shows that the techniques are repeated in the composition of the set, especially for the RF1, SCM1D, and SCPF data sets. The most populous ensembles (RF1, SCM1D, and SCPF data sets) are also the most homoge-

Table 6. Results of first stage experiments on the SWCNT4-Canadija data set with different metrics. The best values are highlighted in dark gray and the second best are in light gray.

Algorithm	MSE	SSE	MAE	R
ARD	32.433369131505	5319.07253756	2.45639544	0.8689465
SGD	40.988686105707	6722.14452133	2.70193416	0.8521731
DT	18.606244702842	3051.42413126	1.70977644	0.9525713
KNN	17.043076586047	2795.06456011	1.57256778	0.9670888
SVM	41.833318058503	6860.66416159	2.65063389	0.5818742
GP	8.877337656136	1455.88337560	1.13649469	0.9829707
RF	13.992511054680	2294.77181296	1.30574379	0.9724443
ET	12.334658531687	2022.88399919	1.26620151	0.9771720
AB	17.476112322384	2866.08242087	2.07287054	0.9238985
GB	10.854874517308	1780.19942083	1.33265800	0.9724447
aASL-MTR	5.423621125583	889.47386459	0.90986227	0.9881044

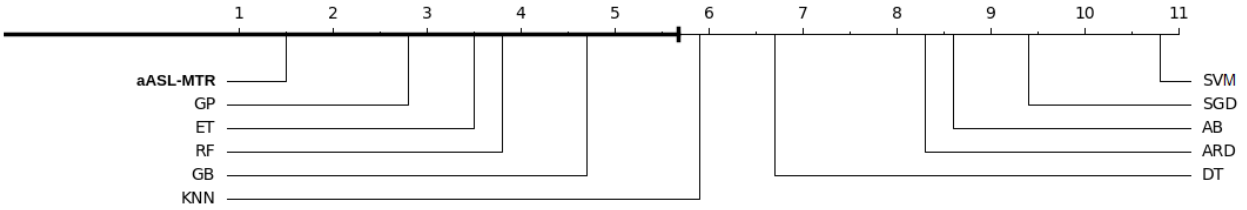

Figure 5. Graphical representation of the Bonferroni-Dunn post-hoc test displaying the average rank of the 11 algorithms considered for the SWCNT4-Canadija data set. The critical difference was equal to 2.813 and aASL-MTR was the baseline algorithm.

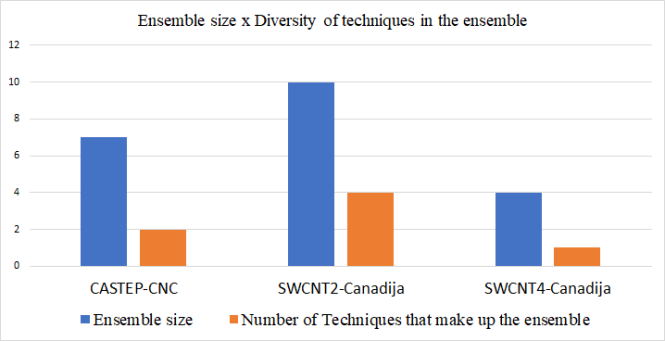

Figure 6. Size and heterogeneity of generated ensembles by aASL-MTR for the carbon nanotubes. The smaller the difference between the orange and blue bars, the greater the heterogeneity of the ensemble.

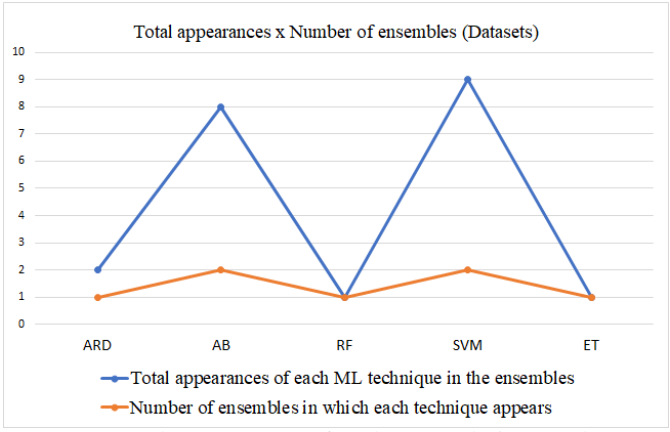

Figure 7. Total appearances of each ML technique and number of ensembles in which each technique appears for the NTC data sets generate by aASL-MTR. The greater the distance between the orange and blue dots, the greater the number of repetitions of the technique the dot represents.

neous. The most heterogeneous ensembles were generated for the JURA, SLUMP, and OSALES data sets. Figure 13 shows the total number of appearances for each technique present in the ensembles of the MTR data sets and also in how many ensembles (data sets) the techniques appear. The techniques that are most repeated in the composition of the ensembles are AB and KNN. The KNN technique is the one that appears the most among the different ensembles (data sets), while the DTs have a total of five appearances in three ensembles (data sets). The SGD and GB techniques do not appear in the ensembles of any of the data sets considered. ET and RF are statistically equivalent to aASL-MTR for the MTR data sets, they appear, respectively, in 8 and 10 ensembles (data sets). ET has in total 16 appearances and RF 14.

5. Conclusions

In this work, we use machine learning techniques, more specifically an AutoML method called Auto Sklearn, which makes

Table 7. Comparison of the aASL-MTR results with the literature for the SWCNT2 and SWCNT4-Canadija data sets. The best values are highlighted in dark gray.

Algorithm	data set	MSE	SSE	MAE	R
aASL-MTR	SWCNT2-Canadija	7.297442022820	1196.78049174	1.06527849	0.9793250
aASL-MTR	SWCNT4-Canadija	5.423621125583	889.47386459	0.90986227	0.9881044
DNN [29]	SWCNT-Canadija	—	—	—	0.9880000

Table 8. Results of second stage experiments on 18 multi-target data sets in terms of aRRMSE metric. The best values are highlighted in dark gray and the second best are in light gray.

data set	ARD	DT	GP	KNN	SGD	SVM	AB	ET	GB	RF	aASL-MTR
ANDRO	0.835	0.432	6.918	0.503	0.795	0.520	0.509	0.522	0.462	0.591	0.351
ATP1D	0.918	1.139	6.283	0.876	4924	1.285	0.780	0.720	0.680	0.693	0.353
ATP7D	1.769	1.317	10.44	1.435	7893	1.409	1.219	1.156	1.276	1.114	0.510
EDM	0.894	0.921	0.750	0.819	0.856	0.718	0.667	0.662	0.831	0.670	0.630
ENB	0.359	0.149	0.169	0.266	0.364	0.302	0.201	0.131	0.065	0.132	0.072
JURA	0.573	0.855	1.246	0.664	0.559	0.566	0.625	0.629	0.664	0.597	0.554
OES10	0.445	0.624	1.267	0.467	2859	0.370	0.449	0.429	0.493	0.437	0.326
OES97	0.604	0.716	1.360	0.574	5762	0.487	0.543	0.538	0.600	0.552	0.493
OSALES	0.876	0.889	1.221	0.897	1.300	0.803	0.945	0.692	0.937	0.757	0.756
RF1	0.718	0.064	0.303	0.127	0.433	0.196	0.281	0.102	0.109	0.077	0.026
RF2	0.369	0.033	0.137	0.031	3923	0.443	0.295	0.020	0.154	0.097	0.034
SCM1D	0.382	0.446	5.180	0.302	8225	0.595	0.461	0.268	0.335	0.291	0.281
SCM20D	0.648	0.570	3.402	0.307	0.652	0.656	0.706	0.428	0.426	0.442	0.307
SCPF	0.990	0.814	2.672	0.895	0.837	0.862	1.287	1.115	1.365	0.888	0.846
SF1	0.988	1.054	1.109	0.996	0.988	1.070	1.100	1.019	1.196	1.035	0.990
SF2	0.942	0.950	1.129	0.951	0.948	1.212	1.120	0.955	1.182	0.947	0.984
SLUMP	0.721	0.832	0.980	0.778	0.731	0.629	0.717	0.659	0.673	0.681	0.587
WQ	0.940	0.941	1.143	0.918	0.940	0.955	0.998	0.893	1.011	0.898	0.906

Table 9. Comparison of the aASL-MTR results with the literature on 18 multi-target data sets in terms of aRRMSE metric. The best values are highlighted in dark gray and the second best are in light gray.

data set	SST [37]	ERC [37]	MMR [39]	SVRCC [38]	MTR-TSF [36]	NOSL [4]	FIRE-ROS [30]	aASL-MTR
ANDRO	0.579	0.567	0.527	0.445	0.360	0.353	—	0.351
ATP1D	0.372	0.372	0.332	0.377	0.362	—	0.413	0.353
ATP7D	0.507	0.512	0.443	0.534	0.423	—	0.523	0.543
EDM	0.740	0.741	0.716	0.697	0.591	0.501	—	0.630
ENB	0.121	0.114	0.111	0.089	0.054	0.080	—	0.072
JURA	0.591	0.590	0.582	0.588	0.569	0.578	—	0.554
OES10	0.421	0.420	0.403	0.353	0.335	—	0.433	0.326
OES97	0.524	0.524	0.497	0.463	0.447	—	0.465	0.493
OSALES	0.726	0.713	0.709	0.781	0.666	—	0.703	0.756
RF1	0.094	0.091	0.089	0.091	0.078	—	0.292	0.026
RF2	0.097	0.095	0.095	0.095	0.068	—	0.291	0.034
SCM1D	0.336	0.330	0.318	0.328	0.299	—	0.384	0.281
SCM20D	0.413	0.394	0.389	0.398	0.301	0.341	0.598	0.307
SCPF	0.831	0.830	0.812	0.801	0.786	—	—	0.846
SF1	1.068	1.089	0.958	0.932	0.779	—	—	0.990
SF2	1.055	1.088	0.984	1.029	0.751	—	—	0.984
SLUMP	0.695	0.689	0.587	0.556	0.528	0.544	—	0.587
WQ	0.909	0.906	0.889	0.904	0.900	—	0.924	0.906

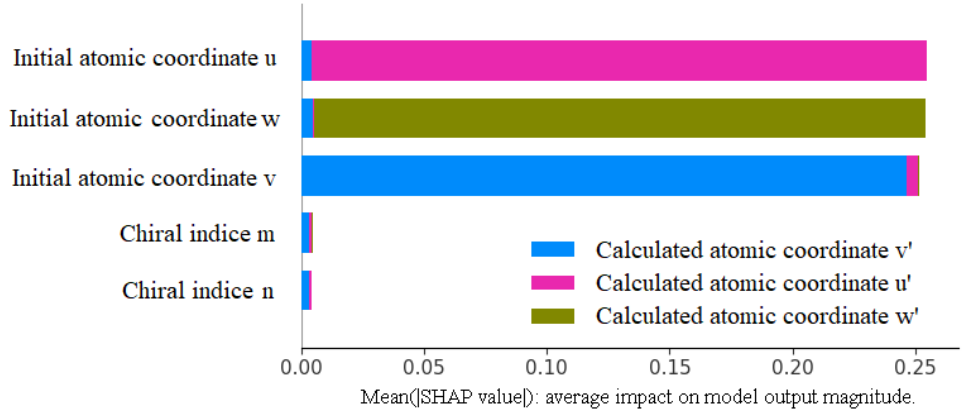

Figure 8. Barplot of 5 attributes from the CASTEP-CNC data set. The X-axis is the mean SHAP values scored by aASL-MTR. The values indicate the average score of the model output for the attributes. Pink, blue and green color represent three different predictions: atomic coordinate v', u' and w', respectively. The Y-axis represents the attributes determined by the explainer.

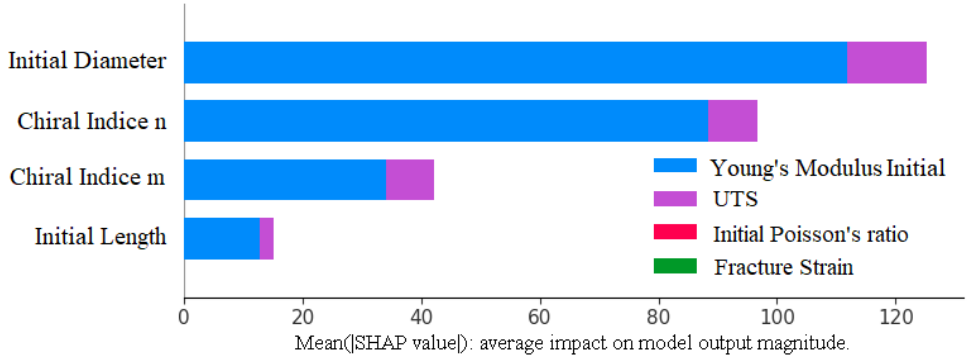

Figure 9. Barplot of 4 attributes from the SWCNT-Canadija data set. The X-axis is the mean SHAP values scored by aASL-MTR. The values indicate the average score of the model output for the attributes. Blue, pink, red and green color represent four different predictions: Young's modulus initial, UTS, Initial Poisson's ratio and Fracture Strain, respectively. The Y-axis represents the attributes determined by the explainer.

use of ensemble learning, being able to select algorithms and their hyperparameter tuning. The Auto Sklearn was adapted and named aASL-MTR. aASL-MTR used a set of 10 ML algorithms (some based on ensemble learning and some not) adjusted/configured for the multi-target regression task.

Two main regression tasks were addressed, one associated with predicting the atomic coordinates of carbon nanotubes and the other with predicting different mechanical properties of carbon nanotubes. In addition, the behavior of the proposal was also observed in another 18 multi-target data sets, encompassing different practical tasks.

In general, aASL-MTR presented good results for all data sets considered, and, compared to the classic algorithms, it was always among the statistically superior algorithms.

Regarding the problems related to carbon nanotubes, aASL-MTR was compared favorably to other algorithms in the literature. It is also concluded in relation to the SWCNT-Canadija data set that using the four attributes of the data set (chiral indeces n and m, initial diameter and initial length) is better than using only two attributes (chiral indeces n and m). The ensembles generated for these tasks vary with respect to the techniques that make up them, the size, and the level of heterogeneity. The AB and SVM techniques stand out in the composition of the ensembles for the CNT data sets. Through SHAP it was possible to identify the importance of attributes for predictions, especially for the SWCNT4-Canadija data set, in which the initial diameter emerged as a high importance attribute.

Considering the remaining 18 data sets for MTR, compared to the classical algorithms, aASL-MTR is the best approach, together with ET and RF. In comparison of the literature, aASL-MTR is among the three best algorithms, which outperform the other techniques compared. The ensembles generated for these data sets are quite diverse. For some data sets, the ensembles are highly heterogeneous and populous, while for some data sets, the ensembles generated are small and homogeneous (or not very diverse). The AB and KNN techniques stand out in the generation of ensembles for MTR data sets.

It should be emphasized that the proposed approach uti-

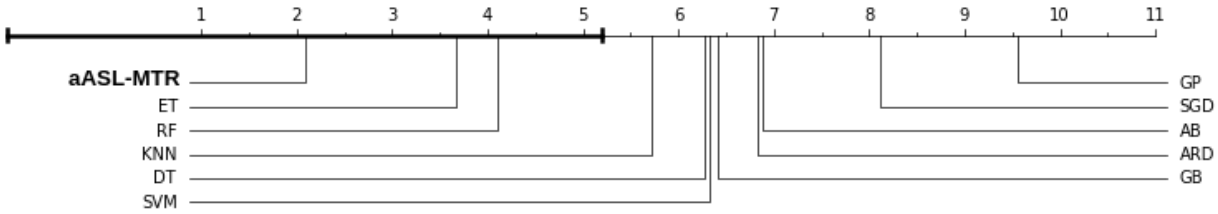

Figure 10. Graphical representation of the Bonferroni-Dunn post-hoc test displaying the aRRMSE average rank for the 11 algorithms considered for the 18 multi-target data sets. The critical difference was equal to 2.813 and aASL-MTR was the baseline algorithm.

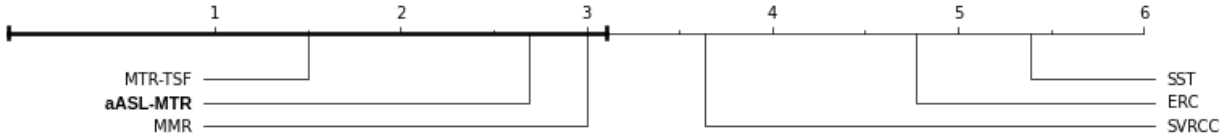

Figure 11. Graphical representation of the Bonferroni-Dunn post-hoc test displaying the average rank of the aRRMSE for the 6 literature algorithms considered for the 18 multi-target data sets. The critical difference was equal to 2.644 and aASL-MTR was the baseline algorithm.

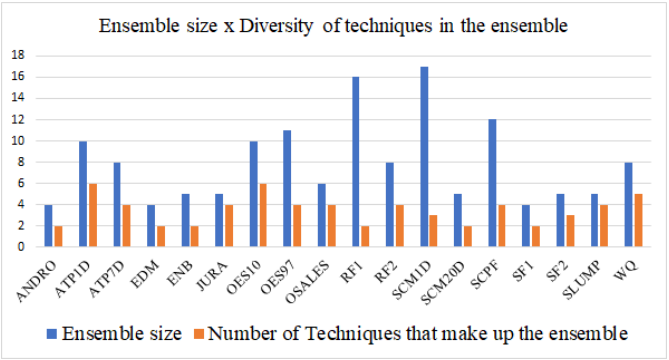

Figure 12. Size and diversity of the ensembles generated by aASL-MTR for the MTR data sets. The smaller the difference between the orange and blue bars, the greater the heterogeneity of the ensemble.

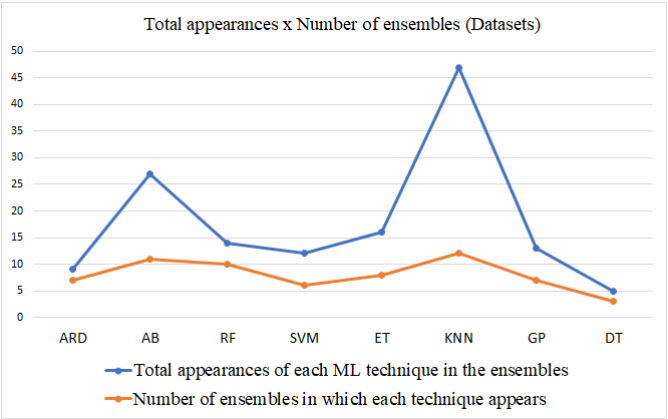

Figure 13. Total appearances of each ML technique and number of ensembles in which each technique appears for the MTR data sets generate by aASL-MTR. The greater the distance between the orange and blue dots, the greater the number of repetitions of the technique the dot represents.

lizes less complex machine learning techniques compared to those discussed in the existing literature. Moreover, aASL-MTR frees users from the burden of configuring the entire ensemble, enabling nonexpert users in the field of machine learning to achieve high-quality models.

In future work, more robust regressors can be included in the AutoML algorithm, such as artificial neural networks, or even use optimization algorithms, such as evolutionary algorithms. In addition, other mechanisms for the automatic construction of ensembles can be tested, as well as other explainable AI tools can be explored. Future research also involves the study and use of statistical tests based on Bayesian optimization during experimental analysis.

Acknowledgments

The authors would like to thank Fundação Araucária for the partial financial support.

Conflict of Interest

The authors declare that they have no conflict of interest.

Author contributions

Authors contribution statement occurred as follows:

Moacir Guedes Oliveira: Formal analysis, Codification, Validation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Supervision;

Sandra M. G. S. Venske: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision;

Carolina P. Almeida: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision.

References

[1] VIVANCO-BENAVIDES, L. E. et al. Machine learning and materials informatics approaches in the analysis of physical properties of carbon nanotubes: A review. *Computational Materials Science*, Elsevier, v. 201, p. 110939, 2022.

[2] RAMAKRISHNA, S. et al. Materials informatics. *J. Intell. Manuf.*, v. 30, n. 6, p. 2307–2326, 2019. Disponível em: ⟨https://doi.org/10.1007/s10845-018-1392-0⟩.

[3] GERON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd. ed. Sebastopol, CA: O'Reilly Media, Inc., 2019. ISBN 1492032646.

[4] ARASHLOO, S. R.; KITTLER, J. Multi-target regression via non-linear output structure learning. *Neurocomputing*, Elsevier, v. 492, p. 572–580, 2022.

[5] HUSMEIER, D. Automatic relevance determination (ard). In: ____. *Neural Networks for Conditional Probability Estimation: Forecasting Beyond Point Predictions*. London: Springer London, 1999. p. 221–227. ISBN 978-1-4471-0847-4. Disponível em: ⟨https://doi.org/10.1007/978-1-4471-0847-4_15⟩.

[6] KETKAR, N. Stochastic gradient descent. In: ____. *Deep Learning with Python: A Hands-on Introduction*. Berkeley, CA: Apress, 2017. p. 113–132. Disponível em: ⟨https://doi.org/10.1007/978-1-4842-2766-4_8⟩.

[7] FÜRNKRANZ, J. Decision tree. In: ____. *Encyclopedia of Machine Learning*. Boston, MA: Springer US, 2010. p. 263–267. ISBN 978-0-387-30164-8. Disponível em: ⟨https://doi.org/10.1007/978-0-387-30164-8_204⟩.

[8] KRAMER, O. K-nearest neighbors. In: ____. *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 13–23. ISBN 978-3-642-38652-7. Disponível em: ⟨https://doi.org/10.1007/978-3-642-38652-7_2⟩.

[9] STEINWART, I.; CHRISTMANN, A. *Support Vector Machines*. Hanover-Pennsylvania: Springer New York, 2008. (Information Science and Statistics). ISBN 9780387772424. Disponível em: ⟨https://books.google.com.br/books?id=HUnqnrpYt4IC⟩.

[10] RASMUSSEN, C. E.; WILLIAMS, C. K. I. *Gaussian processes for machine learning*. Cambridge: MIT Press, 2006. I-XVIII, 1-248 p. (Adaptive computation and machine learning). ISBN 026218253X.

[11] LIU, Y.; WANG, Y.; ZHANG, J. New machine learning algorithm: Random forest. In: LIU, B.; MA, M.; CHANG, J. (Ed.). *Information Computing and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 246–252. ISBN 978-3-642-34062-8.

[12] GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely Randomized Trees. *Machine Learning*, Springer Verlag, v. 36, p. 3–42, 2006. Disponível em: ⟨https://hal.science/hal-00341932⟩.

[13] FREUND, Y.; SCHAPIRE, R. E. Experiments with a new boosting algorithm. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING. *International Conference on Machine Learning*. 1996. p. 148–156. Disponível em: ⟨http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.6252⟩.

[14] FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 29, n. 5, p. 1189 – 1232, 2001. Disponível em: ⟨https://doi.org/10.1214/aos/1013203451⟩.

[15] BRESKVAR, M.; KOCEV, D.; DZEROSKI, S. Ensembles for multi-target regression with random output selections. *Mach. Learn.*, v. 107, n. 11, p. 1673–1709, 2018.

[16] ZHOU, Z.-H. *Ensemble Methods: Foundations and Algorithms*. 1st. ed. Boca Raton, Fla.: Chapman & Hall/CRC, 2012. ISBN 1439830037.

[17] HARRISON, M. *Machine Learning–Guia de Referência Rápida: Trabalhando com dados estruturados em Python*. São Paulo: Novatec Editora, 2019.

[18] RIJN, J. N. van et al. The online performance estimation framework: heterogeneous ensemble learning for data streams. *Machine Learning*, Springer, v. 107, n. 1, p. 149–176, 2018.

[19] HUTTER, F.; KOTTHOFF, L.; VANSCHOREN, J. *Automated Machine Learning: Methods, Systems, Challenges*. 1st. ed. Hanover-Pennsylvania: Springer Publishing Company, Incorporated, 2019. ISBN 3030053172.

[20] DAS, S.; CAKMAK, U. *Hands-On Automated Machine Learning: A beginner's guide to building automated machine learning systems using AutoML and Python*. Birmingham, UK: Packt Publishing, 2018. ISBN 9781788622288. Disponível em: ⟨https://books.google.com.br/books?id=Jd9YDwAAQBAJ⟩.

[21] FEURER, M. et al. Efficient and robust automated machine learning. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS. *Advances in Neural Information Processing Systems 28 (2015)*. [S.l.], 2015. p. 2962–2970.

[22] CARUANA, R. et al. Ensemble selection from libraries of models. In: PROCEEDINGS OF THE TWENTY-FIRST INTERNATIONAL CONFERENCE ON MACHINE LEARNING. *Proceedings of the twenty-first international conference on Machine learning*. [S.l.], 2004. p. 18.

[23] ALI, S. et al. Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, p. 101805, 2023. ISSN 1566-2535. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S1566253523001148⟩.

[24] LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, v. 30, 2017.

[25] ZHANG, C. A.; CHO, S.; VASARHELYI, M. Explainable artificial intelligence (xai) in auditing. *International Journal of Accounting Information Systems*, Elsevier, v. 46, p. 100572, 2022.

[26] ACI, M.; AVCI, M. Artificial neural network approach for atomic coordinate prediction of carbon nanotubes. *Applied Physics A: Materials Science and Processing*, v. 122, n. 7, 6 2016. Disponível em: ⟨https://doi.org/10.1007/s00339-016-0153-1⟩.

[27] SAUL, C. J.; KAMIS, A. Z.; SAHIN, E. Carbon nanotube coordinate prediction with deep learning. In: 4TH INTERNATIONAL CONFERENCE ON COMPUTER AND COMMUNICATION SYSTEMS (ICCCS). *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*. [S.l.], 2019. p. 1–6.

[28] ATES, M.; EKER, A. A.; EKER, B. Carbon nanotube-based nanocomposites and their applications. *Journal of Adhesion Science and Technology*, Taylor & Francis, v. 31, n. 18, p. 1977–1997, 2017. Disponível em: ⟨https://doi.org/10.1080/01694243.2017.1295625⟩.

[29] CANADIJA, M. Deep learning framework for carbon nanotubes: Mechanical properties and modeling strategies. *Carbon*, Elsevier, v. 184, p. 891–901, 2021.

[30] BRESKVAR, M.; DZEROSKI, S. Multi-target regression rules with random output selections. *IEEE Access*, v. 9, p. 10509–10522, 2021.

[31] LARGE, J.; LINES, J.; BAGNALL, A. A probabilistic classifier ensemble weighting scheme based on cross-validated accuracy estimates. *Data Min. Knowl. Discov.*, Kluwer Academic Publishers, USA, v. 33, n. 6, p. 1674–1709, nov 2019. ISSN 1384-5810. Disponível em: ⟨https://doi.org/10.1007/s10618-019-00638-y⟩.

[32] XAVIER-JúNIOR, J. C. et al. An evolutionary algorithm for automated machine learning focusing on classifier ensembles: An improved algorithm and extended results. *Theoretical Computer Science*, v. 805, p. 1–18, 2020. ISSN 0304-3975. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S0304397519307649⟩.

[33] ZHANG, X.; MAHADEVAN, S. Ensemble machine learning models for aviation incident risk prediction. *Decision Support Systems*, v. 116, p. 48–63, 2019. ISSN 0167-9236. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S0167923618301660⟩.

[34] DUMANCAS, G.; ADRIANTO, I. A stacked regression ensemble approach for the quantitative determination of biomass feedstock compositions using near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, Elsevier, v. 276, p. 121231, 2022.

[35] CARNEIRO, T. C. et al. Ridge regression ensemble of machine learning models applied to solar and wind forecasting in brazil and spain. *Applied Energy*, Elsevier, v. 314, p. 118936, 2022.

[36] WANG, J. et al. Multi-target regression via target specific features. *Knowledge-Based Systems*, Elsevier, v. 170, p. 70–78, 2019.

[37] SPYROMITROS-XIOUFIS, E. et al. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, Springer, v. 104, n. 1, p. 55–98, 2016.

[38] MELKI, G. et al. Multi-target support vector regression via correlation regressor chains. *Information Sciences*, Elsevier, v. 415, p. 53–69, 2017.

[39] ZHEN, X. et al. Multi-target regression via robust low-rank learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 40, n. 2, p. 497–504, 2017.

[40] FACELI, K. et al. *Inteligência artificial: uma abordagem de aprendizado de máquina*. [S.l.]: Grupo Gen - LTC, 2011.

[41] MÜLLER, A. C.; GUIDO, S. *Introduction to machine learning with Python: a guide for data scientists*. Sebastopol, CA: "O'Reilly Media, Inc.", 2016.

[42] JASTRZEBSKI, Z. *Nature and Properties of Engineering Materials*. Hoboken, NJ: Wiley, 1959. (Wiley international edition). ISBN 9780471440888. Disponível em: ⟨https://books.google.com.br/books?id=ucY3zQEACAAJ⟩.

[43] HALL, L. J. et al. Sign change of poisson's ratio for carbon nanotube sheets. *Science*, v. 320, n. 5875, p. 504–507, 2008. Disponível em: ⟨https://www.science.org/doi/abs/10.1126/science.1149815⟩.

[44] SCAFFARO, R.; BOTTA, L. Chapter 5 - nanofilled thermoplastic–thermoplastic polymer blends. In: THOMAS, S.; SHANKS, R.; CHANDRASEKHARAKURUP, S. (Ed.). *Nanostructured Polymer Blends*. Oxford: William Andrew Publishing, 2014. p. 133–160. ISBN 978-1-4557-3159-6. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/B9781455731596000055⟩.

[45] HATZIKOS, E. V. et al. An empirical study on sea water quality prediction. *Knowledge-Based Systems*, Elsevier, v. 21, n. 6, p. 471–478, 2008.

[46] GROVES, W.; GINI, M. On optimizing airline ticket purchase timing. *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM New York, NY, USA, v. 7, n. 1, p. 1–28, 2015.

[47] KARALIČ, A.; BRATKO, I. First order regression. *Machine learning*, Springer, v. 26, n. 2, p. 147–176, 1997.

[48] TSANAS, A.; XIFARA, A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, Elsevier, v. 49, p. 560–567, 2012.

[49] GOOVAERTS, P. *Geostatistics for Natural Resources Evaluation*. Northamptonshire: Oxford University Press, 1997. (Applied geostatistics series). ISBN 9780195115383. Disponível em: ⟨https://books.google.com.br/books?id=CW-7tHAaVR0C⟩.

[50] CHUDZICKI, D. *Online product sales*. 2012. Online. Disponível em: ⟨https://www.kaggle.com/c/online-sales⟩.

[51] CUKIERSKI, W. *See Click Predict Fix*. 2013. Online. Disponível em: ⟨https://www.kaggle.com/c/see-click-predict-fix⟩.

[52] BACHE, K.; LICHMAN, M. *UCI Machine Learning Repository*. 2013. Online. Disponível em: ⟨http://archive.ics.uci.edu/ml⟩.

[53] YEH, I.-C. Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cement and concrete composites*, Elsevier, v. 29, n. 6, p. 474–480, 2007.

[54] DŽEROSKI, S.; DEMŠAR, D.; GRBOVIĆ, J. Predicting chemical parameters of river water quality from bioindicator data. *Applied Intelligence*, Springer, v. 13, n. 1, p. 7–17, 2000.

[55] SHAHHOSSEINI, M.; HU, G.; PHAM, H. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Machine Learning with Applications*, Elsevier, v. 7, p. 100251, 2022.

[56] DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, JMLR. org, v. 7, p. 1–30, 2006.

[57] DUNN, O. J. Multiple comparisons among means. *Journal of the American Statistical Association*, Taylor & Francis, v. 56, n. 293, p. 52–64, 1961.